An acoustical study of Korean monophthongs produced by male and female speakers

Byunggon Yang
Department of English, College of Humanities, Dongeui University, 24 Kaya-dong, Pusanjin-gu, Pusan 614-714, Korea

(Received 21 August 1990; accepted for publication 29 November 1991)
The first three formants of the ten Korean monophthongs were studied. The vowels were produced by 20 male and female speakers in an \( hVd \) (\( V = \) vowel) context. Then the male-female variations in the Korean formant data were examined. The uniform scaling method resulted in less than a 6\% difference between the scaled and reference data. A regression analysis showed that the slope of a regression equation with a small intercept near the origin could be used for the uniform scale factor.

INTRODUCTION

Previous studies have shown that formant frequencies are important factors in determining the perceptual quality of a vowel. Hence, each vowel phoneme has its characteristic formant patterns within a given language. However, spectrographic studies of vowels produced by different speakers of a given language show great variation in formant frequencies dependent on age and gender. This speaker variation causes many problems in determining the acoustic correlates of the vowel qualities of a language. These sources of speaker variation are divided into linguistic and nonlinguistic factors (Traunmüller, 1988). Linguistic factors are dialectal, sociolectal, idiolectal, and phonostylistic differences. Nonlinguistic factors include the differences in age, gender, and the emotional state of the speakers. Since the vowel formants depend on the length and shape of the vocal tract, a major source of nonlinguistic speaker differences is the individual speaker's vocal tract anatomy (Fant, 1960, 1968; Nordström and Lindblom, 1975; Nordström, 1975).
The goal of factoring out these nonlinguistic factors is to establish a "pure," linguistically relevant acoustic specifica-
tion of the vowel qualities of any given language. This procedure has been called "normalization" (Fant, 1968). Various proposals for the normalization of vowel qualities have been explored, but the problem has not yet received its definitive solution. Some proposals are articulatorily based (Nordström and Lindblom, 1975; Fant, 1975), while others are auditorily based (Miller et al., 1980; Miller, 1989; Syrdal and Gopal, 1989), involving the transformation of the formant values into auditory units. Auditorily normalizations are based on the belief that there is a constancy in the spatial pattern of excitation in the peripheral auditory system for a given vowel produced by different speakers. In this study the discussion is limited to the two articulatorily based proposals: uniform and nonuniform scaling methods. Nordström and Lindblom (1975) proposed a uniform scaling based on an estimate of the length of the speaker's vocal tract. According to their claim, all observed speaker-dependent formant frequency differences can be satisfactorily accounted for solely in terms of vocal tract length. Assuming that overall vocal tract length variations affect all formant frequency values by the same scale factor (that is, in a uniform way), the same scale factor was applied to all formants. Therefore, the term "uniform scaling" was used. It involved estimating the total length of a subject's vocal tract from an average of F3 in vowels with F1 greater than 600 Hz. Because the length of the speaker's vocal tract is inversely related to formant frequency, the ratio of the length of the average male vocal tract (Lm) to the average female vocal tract length (Lf) can be written:

\[ k = \frac{L_f}{L_m} = \frac{F_{3,m}}{F_{3,f}} \]  

(1)

in which \( F_{3,m} \) and \( F_{3,f} \) indicate an average of the third male and female formant values, respectively. Then, the normalized nth female formant frequency is denoted as \( F_{nf} \) (scaled) and can be determined according to

\[ F_{nf}(scaled) = kF_{nf} \]  

(2)

On the other hand, Fant (1968) observed that female speakers have proportionately shorter pharynges and argued that the scaling of formant values must be nonuniform. In other words, it must consider not only differences in the
overall vocal tract length between male and female speakers but also the complex formant-cavity relationships. Therefore, Fant (1975) recommended using scale factors that are both vowel and formant specific. His method applies different scale factor to each individual vowel and individual formant category. A tentative reference table of "universal" scale factors was proposed based on the statistical study of six languages: Swedish, American English, Danish, Estonian, Dutch, and Serbo-Croatian (Fant, 1975, p.5, Table I-A-I). Because the table was derived from the average differences of male and female formant frequencies only for those six languages, it may not be proper to use those factors in normalizing the Korean data.

In this paper, the first three formants of the ten Korean monophthongs produced by 20 male and female speakers were studied while controlling the linguistic factors as homogeneously as possible in each group. Second, the male-female variations in the Korean data were examined. Third, the data were scaled using the uniform scaling method. Fourth, regressive analyses were conducted on formant values of male and female groups.

I. METHOD
A. Subjects and speech samples
A total of 20 subjects were chosen from a larger group participating in recording and listening sessions at the University of Texas at Austin. They formed two groups: ten Korean males and ten Korean females. Subjects were students attending the University of Texas at Austin and all had normal hearing and health. All the Korean subjects spoke Standard Korean. Two screening instruments were used to make each group linguistically homogeneous.

First, subjects were grouped homogeneously on the basis of collected information from a questionnaire. The questionnaire included subjects’ name, age, sex, height, native language, fluency in other languages, dialect, and history of speech and hearing disorders. Second, peer judgment was employed to screen
out those subjects who had different dialects. Five peers in each male and female group were randomly chosen from among the subjects. Then the ten peers were asked to listen to four sets of tokens consisting of the vowels [i e a u]. Each set is composed of different male and female subjects saying the same token. In the listening session, the peers put a check mark on each token that sounded different from their own dialect. All the marks were counted to find four peers (two males and two females) who had the fewest marks. Finally, marks by the four peers were used to reject those subjects who had more than 80% of the total tokens perceived as a different dialect.

The speech samples consisted of 52 Korean words. Each Korean vowel occurred in an [h(V)d] (V = vowel) context. In this context, the following vowel formant can be easily identified because the [h] noise on the spectrogram shows similar patterns of the following vowel formants. Ten Standard Korean monophthongs studied were /a e e o o u u i i i/. These ten Korean vowels appeared five times in random order. Later, three out of the five productions of each vowel were randomly chosen for the average data set, avoiding unnatural tokens at the beginning and ending of the recording.

B. Procedures

The recording was done in a sound-proof booth in the Phonetics Lab of the University of Texas at Austin (UT). The experimenter asked the subjects to produce each word at a normal rate and as naturally as possible. Speech samples were picked up by a condenser microphone at a distance of about 15 cm. A professional cassette recorder (Model CD-330) was used for recordings. The recording level was carefully monitored by the experimenter throughout the recording session to avoid ultraweak or overloaded signals. The recording took 2-3 min per subject.

The recorded samples were analyzed using the VAX computer in the UT Phonetics Lab. The KLSPEC software package written by D. H. Klatt and modified by J. Lame was used to interactively examine, measure, and analyze the recorded samples. The samples were played from a Yamaha (model KX-800U) cassette deck. The input samples were low pass filtered at 4 kHz and digitized at a 10-kHz sampling rate. Spectrograms were made using a 256-point discrete
Fourier transform (DFT) analysis with a 6.4-ms Hamming window once every millisecond. The dynamics of the vowel formant pattern made it difficult to find a consistent time point for spectrum analysis. Therefore, the author calculated the total duration from the vowel onset and offset time points. Each time point was determined by adding one-third of the total duration to the vowel onset point.

C. Formant decisions

In this study, a spectrogram of each vowel was used for the first reference and final decision on formant values. First, formant values on the spectrogram were estimated by drawing a pencil line through the center of each formant band with a ruler. Then, visual estimates were compared with two DFT harmonic spectra (average envelope and LPC envelope) estimates. The output of the software package listed estimates of three or more feasible formant values along with amplitude levels. These methods almost always converged. However, when formant values of the same vowel for a subject showed a difference of more than 500 Hz, they were rejected under the assumption that the subject produced a different target vowel. For reliability, the measurements were checked by an independent observer. There was an agreement of more than 97% in the formant decision.

II. RESULTS AND DISCUSSION

The average values of the first three formants (F1, F2, and F3) and their standard deviation (s.d.) of the Korean monophthongs are listed in Tables I and II. As is shown in the tables, the deviation within males is generally lower than that of females. Figure 1 illustrates the vowel space of males and females in which adjacent vowel points are connected peripherally. Phonemes are given near various symbols. The

<table>
<thead>
<tr>
<th>Vowel</th>
<th>F1</th>
<th>s.d.F1</th>
<th>F2</th>
<th>s.d.F2</th>
<th>F3</th>
<th>s.d.F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>986</td>
<td>107</td>
<td>1794</td>
<td>108</td>
<td>2957</td>
<td>227</td>
</tr>
<tr>
<td>e</td>
<td>677</td>
<td>108</td>
<td>2285</td>
<td>169</td>
<td>3063</td>
<td>141</td>
</tr>
<tr>
<td>e</td>
<td>650</td>
<td>113</td>
<td>2377</td>
<td>77</td>
<td>3068</td>
<td>117</td>
</tr>
<tr>
<td>i</td>
<td>344</td>
<td>48</td>
<td>2814</td>
<td>168</td>
<td>3471</td>
<td>177</td>
</tr>
<tr>
<td>o</td>
<td>499</td>
<td>60</td>
<td>1029</td>
<td>143</td>
<td>3068</td>
<td>159</td>
</tr>
</tbody>
</table>

TABLE I. Average values of the first three formants (F1, F2, F3) and their standard deviation (s.d.F1, s.d.F2, s.d.F3) of the Korean female speakers.
TABLE II Average values of the first three formants (F1,F2,F3) and their standard deviation (s.d.F1, s.d.F2, s.d.F3) of the Korean male speakers.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>F1</th>
<th>s.d.F1</th>
<th>F2</th>
<th>s.d.F2</th>
<th>F3</th>
<th>s.d.F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>738</td>
<td>81</td>
<td>1372</td>
<td>124</td>
<td>2573</td>
<td>127</td>
</tr>
<tr>
<td>e</td>
<td>591</td>
<td>75</td>
<td>1849</td>
<td>106</td>
<td>2597</td>
<td>110</td>
</tr>
<tr>
<td>e</td>
<td>490</td>
<td>105</td>
<td>1968</td>
<td>150</td>
<td>2644</td>
<td>94</td>
</tr>
<tr>
<td>i</td>
<td>341</td>
<td>29</td>
<td>2219</td>
<td>176</td>
<td>3047</td>
<td>146</td>
</tr>
<tr>
<td>o</td>
<td>453</td>
<td>47</td>
<td>945</td>
<td>134</td>
<td>2674</td>
<td>156</td>
</tr>
<tr>
<td>o</td>
<td>459</td>
<td>69</td>
<td>1817</td>
<td>163</td>
<td>2468</td>
<td>134</td>
</tr>
<tr>
<td>u</td>
<td>369</td>
<td>43</td>
<td>981</td>
<td>141</td>
<td>2565</td>
<td>173</td>
</tr>
<tr>
<td>ü</td>
<td>338</td>
<td>30</td>
<td>2114</td>
<td>140</td>
<td>2729</td>
<td>213</td>
</tr>
<tr>
<td>ʌ</td>
<td>608</td>
<td>76</td>
<td>1121</td>
<td>110</td>
<td>2683</td>
<td>145</td>
</tr>
<tr>
<td>i</td>
<td>405</td>
<td>37</td>
<td>1488</td>
<td>176</td>
<td>2497</td>
<td>80</td>
</tr>
</tbody>
</table>

same symbol indicates the same phoneme. The two vowel spaces appear triangular. A thicker line connects male vowels, while a thinner line connects female vowels. The vowel spaces show some systematic relationship. It implies a trend toward higher formant frequencies for female speakers. This tendency may be largely due to nonlinguistic factors because linguistic factors in this experiment were homogeneously controlled in each group as much as possible. To examine that relationship, the percent difference (Diff.) in the male and female formant frequencies across vowel formant was calculated by

\[
\text{Diff.}(\%) = \frac{(F_{nf} - F_{nm})}{F_{nm}} \times 100, \quad (3)
\]

in which \(F_{nf}\) denotes the \(n\)th female formant value. The average Diff. across all the vowels comes out to 18% which equals the average difference of F1. In F2 and F3 the Diffs. are 20% and 17%, respectively. In addition the greatest variation within each formant was in F1 (s.d. = 12%) fol-
lowed by F2 (s.d.= 9%) and F3 (s.d.= 3%). This implies that male data may be uniformly shifted to the female data or vice versa. Thus the uniform scaling method was applied to the Korean data. The uniform scale factor is 0.87 estimated from Eq. (1). This ratio corroborates the study by Chiba and Kajiyama (1941) who estimated overall vocal tract length assigning the relative numbers of 1.0 to males and 0.87 to females. The data were scaled by Eq. (2). Then, a numerical criterion was used to see how closely the female data were scaled to the male reference data (Fant, 1975). For that purpose, Fnf in Eq. (3) was replaced by Fnf (scaled). The uniform scaling method had an average Diff. of less than 6% across all the vowels. Specifically, there was more than a 10% Diff. in F1 of the uniformly-scaled data, but less than a 5% Diff. was observed in back vowels. In F2 the point vowels [i a u] showed a 10% Diff. while the others were less than 10%. In the vowel /a/ a perfect match occurred. In F3t an average Diff. of 2% was observed which stands to reason since the scale factor was derived from the ratio of male to female averages of F3.

The question arises as to whether the nonuniform scaling greatly improves the uniform scaling method in the Korean data. Fant (1975) suggested that the gender variation can be best resolved by formant-number and vowel-specific scale factors. Certain Korean vowels show greater male-female difference. For example, F1 in the vowel [0] shows a
34% Diff. while that of the vowel [i] shows less than a 1% Diff. Since the scale factors in the nonuniform methods were derived from the average variation of the six European languages they may not be, proper to use in the scaling of the Korean data. However, one can reason that if one employs a formant-number-specific method, the average Diff. would improve around 1% because there was about 1% difference among F1, F2, and F3 in the previous analysis. Moreover, if both the formant-number and vowel-specific scale factors are used, then it will improve less than 6%.

Finally, regressive analyses were conducted to find regression equations or the best-fitting lines for the relationship between all the female formant values (Fnf) and those of males (Fnm) or vice versa:

\[
\begin{align*}
F_{nm} &= 0.85 F_{nf} - 5 \\
F_{nf} &= 1.18 F_{nm} + 19 \quad (r^2 / 0.99)
\end{align*}
\]

The slope 0.85 is similar to the average of F3/f/ F3/m ( \( r^2 = 0.87 \) ) for the uniform scaling method. With small intercepts, both equations imply that one may expect a good fit with a regression equation through the origin (zero intercept). Thus the uniform scale factor can be easily determined through the regression. The \( r^2 \) indicates that female formant values can be accurately predicted from male values or vice versa. In addition, a regression analysis between female formant frequency and the male-female difference in percent revealed a strong correlation (Fnf against Diff., \( r < 0.78 \) ). This corroborates Fant's observation (1975) that the gender difference in F1 and F2 increases with formant frequency. O'Leary (1989) has also observed a strong correlation between scale factors obtained with Fant's scaling method and female formant values. Fant's scale factors were higher than the average for high vowels while those for lower vowels were lower.

Using the regression equations relating Fant's scale factors to female formant frequencies, O'Leary scaled female formant frequencies to match more closely those of males.

In conclusion, the first three formants of the ten Korean monophthongs were studied. Because the linguistic factors were controlled to be as homogeneously as possible in each group, the male-female difference was attributed to nonlinguistic factors. The uniform scaling method resulted in less than a 6% difference between the scaled and reference data. A regression analysis showed that the slope of a regression
equation with a small intercept near the origin could be used for the uniform scale factor.


